

**Original citation:**

Michael, John and Wolf, Thomas. (2014) Why apply a hierarchical predictive processing framework to music perception and performance? *Empirical Musicology Review*, 9 (3-4). pp. 177-182.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/80915>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0) license and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by-nc/4.0/>

**A note on versions:**

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# Why Apply a Hierarchical Predictive Processing Framework to Music Perception and Performance?

JOHN MICHAEL[1]

*Central European University*

THOMAS WOLF

*Central European University*

**ABSTRACT:** In this commentary, we examine Schaefer's proposal to apply a hierarchical predictive processing (HPP) framework to research on music perception and music performance. As we shall see, this proposal raises the possibility of enriching this research area with new theoretical and empirical resources from further afield. In order to tap the potential of such a project, it will be important to work towards the formulation of hypotheses that are uniquely generated by the HPP framework. We attempt to contribute to this project by specifying explanatory resources within the HPP framework that are well-suited to formulating such hypotheses, by articulating several novel questions that are generated by the attempt to apply the HPP framework to musical perception and performance, and by identifying potential challenges for such a project to address. We also provide reasons to expect that the HPP framework may be especially fruitful in the context of joint (musical) action.

Submitted 2014 December 11; accepted 2014 December 23.

**KEYWORDS:** *music perception, predictive coding, embodied cognition, joint action*

IN an upcoming article in this special issue, Schaefer (2014) demonstrates how a hierarchical predictive processing (HPP) framework can be drawn upon to create an illuminating new perspective upon music perception and music performance. Such a perspective makes it possible to give fresh answers to old questions, to establish novel connections among apparently disparate themes and research areas, and to raise stimulating new questions for further research. Moreover, as Schaefer emphasizes, the framework is also appealing for broad theoretical reasons – in particular, it is well-positioned to integrate core insights animating embodied cognition approaches (such as the interconnectedness of cognition, action and perception), while at the same time retaining a focus on the internal models that enable music listeners and performers to anticipate and to synchronize with musical patterns, and to coordinate with others in producing music. In this brief commentary, we will home in on a few of the central themes of Schaefer's article, and attempt to identify several ways in which those themes may fruitfully be articulated further. Let's begin by recalling the basic outline of Schaefer's approach.

## MUSIC IN THE PREDICTIVE BRAIN

At the core of the HPP framework is the idea of a hierarchy of models (for comprehensive accounts, see Hohwy 2013; Clark 2013; Friston 2012). Lower-level models predict changes in sensory input at short timescales, while higher-level models predict changes occurring over longer timescales, namely by representing more abstract causes of the changes in sensory input. Models at hierarchically different levels constrain each other through the dynamics of prediction error minimization: each model generates predictions about the representations at the immediately subordinate level; these predictions are compared with the actual state of the subordinate-level model; a prediction error is returned to the superordinate-level model; the superordinate-level model is then revised to generate a new prediction. By this process, the interconnected models are continuously updated and prediction errors minimized.

In the context of music, such a hierarchy of models makes it possible to predict musical events at multiple timescales, from predictions of events occurring over shorter timescales (such as what the next tone will be) to slower changes at hierarchically superordinate levels (such as chord progressions and upcoming resolutions). Even higher up the hierarchy, there may also be beliefs about, for example, the performer's or the composer's intentions, or the social context of the music. An important question to ask, moving forward, is what specific hypotheses are generated by this conception of a vertical hierarchy of models, organized around timescales and degrees of abstraction. Psychologists interested in music perception have been alert to the importance of anticipation for many years (Keller, Novembre & Hove 2014), so the emphasis on prediction (i.e. anticipation) is not itself new. In order to generate truly novel hypotheses, it seems most promising to examine the specific dynamics of prediction error minimization across the vertical hierarchy, i.e. the interplay of predictions generated at different levels of the cortical hierarchy. How might it be possible, for example, to modulate music listeners' expectations about events occurring over shorter timescales by manipulating their background beliefs about the music? In other areas of cognitive science, such as in research on visual perception, HPP has been used to generate novel questions and hypotheses, and to formulate specific quantitative models. It may be fruitful to see how this has been done in those contexts, and to look for analogous phenomena in the context of music perception or performance. In one pioneering study, for example, Hohwy and colleagues (2008) were able to use HPP to formulate highly specific predictions about top-down effects upon the temporal pattern of changes in visual perception within a binocular rivalry paradigm. Might it be possible to create effects akin to binocular rivalry in the context of music perception – for example by creating stimulus configurations that can be heard as manifesting two separate rhythms, but not both at the same time? For example, in a stream segregation paradigm, participants might hear a target rhythm in one segregated stream or a different rhythm if they hear the two streams as unsegregated (Bregman, 1994). If this could be done in a way that leads listeners to switch periodically between hearing the two rhythms, then it would be interesting to investigate whether manipulating their high-level expectations (for example by manipulating their background beliefs or by instructing them beforehand to move their bodies with one of the rhythms) may influence the pattern of shifts in their perception of the rhythm.

Ongoing work by Wolf, Vesper, Knoblich, and Sebanz (2014) on the top-down influence of beliefs upon music perception and performance in a joint action paradigm may provide a starting point for further research along these lines. In this study, novices performed very simple piano duets in dyads. For half of the duets, participants were informed that the two parts had been composed with the intention to be performed as duets, whereas for the other half of the duets, they were informed that the two parts had not been composed with this intention. In addition, Wolf and colleagues manipulated the harmonic relation between the two parts (same key or different keys). The main finding was that the differences in synchrony between the same key and the different key parts were significantly greater when participants believed the melodies had been intended to be played together, attesting that the participants' beliefs had led to an increase of joint action representations.

## PREFERENCES AND PREDICTABILITY

One important use to which Schaefer wants to put this hierarchical apparatus is in accounting for individual differences in music perception, in particular individual differences in *preferences*. It is clear enough that people would come to have different models based upon their different music experiences, and that this would lead them to have different expectations, but we do not find it obvious how this would lead to differences in preference. In order to explain how differences in preference arise, it appears necessary to introduce further explanatory resources. It would be clear, for example, that different predictive models lead to different preferences if it were the case that people have a preference for what is expected. But *is* this the case? Too much predictability can of course also be boring. One possible solution, then, is that there is a sweet spot of predictability, such that each person tends to prefer music that is just predictable enough to engage with their models, but not so predictable as to be boring – much as Goldilocks always opts for the middle ground between the hard bed and the soft bed, between the hot porridge and the cold porridge. Huron (2006), for example, has developed an account of emotional valence in musical perception based upon this idea. As he puts it: “Expectations that prove to be correct represent successful mental functioning. Successful predictions are rewarded by the brain” (p. 361); and, on the other hand: “Although expected events are generally preferred, highly predictable environments can lead to reduced attention and lowered arousal – often leading to sleepiness” (p. 362) More generally, accounts of aesthetic experience

have also been developed based upon this idea, which is often referred to under the label of “processing fluency” (Reber, Schwarz and Winkielman, 2004; Berlyne 1971).

But this idea does not seem plausible. First, the amount of predictability that people prefer seems to be among the things that differ from one person to the next: some like repetitive music in the style of Philip Glass, while some prefer free jazz (Rentfrow & Gosling, 2003; Francès, 1976; Kammann, 1966; Barron, 1955). Moreover, even if there were a sweet spot for predictability, we do not think that identifying that sweet spot could be sufficient to account for preferences. This is because there are qualitative differences among the (tonal) events that are being predicted in the first place. Highly unpleasant sounds, such as those produced by a novice trumpet player, may not be very congenial even if they have the “right” degree of predictability. Thus, we suspect that the notion of a sweet spot of predictability may in fact merely create the impression of an explanation without actually delivering one (we might dub this the “Goldilocks fallacy”). In sum, we think that one important challenge moving forward will be to give a detailed account of how different predictive models lead to different preferences.

### EMBODYING MUSIC

A further attractive feature of the HPP framework, as Schaefer emphasizes, is that it is well-positioned to capture and do justice to the insights of embodied cognition approaches without embracing the skepticism about representations and internal models which such approaches have sometimes led to, and thus without sacrificing the explanatory power of such internal models (De Bruin & Michael, forthcoming; Geeves, McIlwain, Sutton & Christensen, 2013; Michael, Christensen & Overgaard, 2014). This is because, like embodied cognition approaches, HPP places an emphasis on the integration of cognition with both perception and action. With respect to perception, it denies that there is any clear divide between perception and cognition. Instead, according to HPP, there is a graded hierarchy of models functioning according to the same basic principles but differing with respect to the timescale at which they generate predictions of changes in the world. With respect to action, as Schaefer also points out, the HPP framework in fact identifies a sense in which action can be *equivalent* to the updating of higher-level cognitive models, namely through active inference. The idea is as follows. When confronted with a prediction error, the brain has two options. The first option is to revise its model of the world until the prediction error is satisfactorily diminished. The second option is to change the world so that it matches the model. In the context of music performance, for example, one may expect to hear the sounds one is producing on a piano at a particular tempo; if this expectation is not met, one can revise one’s expectations or one can actively change the tempo. In the terms of HPP, this would amount to changing the world to match the model one had of it.

In future research, it will be interesting to investigate whether this novel view of how action relates to perception and cognition leads to new questions and to novel and testable hypotheses. The notion of active inference may, for example, create a fresh perspective upon the phenomenon of rhythmic synchronization, i.e. on people’s tendency to synchronize their bodily movements involuntarily with perceived rhythmic tones or light emissions (see Repp 2005). One might, for example, conceptualize this tendency as a response to the prediction errors that are generated as the brain perceives (and attempts to model and predict) rhythmic tones. In other words, perhaps the brain engages in active inference and expects thereby to influence the perceived rhythm and to make it conform to the model. This would only be the case if the brain had a prior expectation that bodily movement is likely to influence perceived rhythms, or that perceived rhythms are coupled to moving bodies – perhaps because this was the case in many contexts evolutionarily. This raises questions about the kinds of rhythmic tones or light emissions the brain might have modeled and or influenced through bodily movements in what evolutionary contexts. Although speculations of this nature about evolutionary history are notoriously difficult to evaluate, this line of thought does illustrate the kind of conjecture that the notion of active inference can inspire. Alternatively, one might cast the same conjecture in ontogenetic terms: perhaps the brain expects to be able to influence the rhythm of perceived auditory or visual stimuli because it often *can* during the course of development. This is not altogether unlikely, given the importance of contingent and rhythmic social interaction in early childhood (Gergely & Watson, 1999). This line of thought leads to the prediction that people might have a bias toward expecting that their own rhythmic movements will influence the rhythm of perceived tones or light emissions.

## PLAYING AND LISTENING TOGETHER: JOINT ACTION IN MUSIC

One further area in which the integration of action with perception and cognition may be especially fruitful is the context of *joint* musical action. When the perspective is expanded from individual agents listening to or performing music to multiple agents doing so together, several important new factors come into the picture. First of all, each agent's performance of or response to the music can influence the other agents' performance and response. Thus, there is a straightforward application for the notion of active inference. Second, the reciprocal influence among the different agents also occurs at the level of the models they employ to predict what is coming next in the music. This is because each agent is not only modeling the music but is also modeling the other agent's actions and intentions, as well as the other agent's model of her actions and intentions. As Hohwy and Palmer (2014; see also Friston & Frith, in press) have pointed out, this kind of recursive higher-order modeling is characteristic of human social interaction, and generates new, characteristically social challenges for the brain engaged in prediction error minimization. For example, it casts joint musical performance in the form of a coordination problem. In coordination problems, such as the stag hunt game (Skyrms, 2004), two agents each maximize their individual benefits if they coordinate their actions but get no benefit if they try to coordinate but fail to do so, and are therefore tempted to opt for a smaller benefit which does not depend upon coordinating with the other agent. In the context of music, two jazz performers might both prefer, having just finished what appeared to be the final verse of a jazz standard, to start up again for a surprising additional repetition of the chorus – *but only if the other does so as well*. In this case, the structure of a coordination problem is implemented, since the agents' interests converge, but there is also a second acceptable option (remaining silent), which entails less risk. Each agent has to assess what the other agent is likely to expect her to do, what that other agent is likely to expect her to expect the other agent to, etc. In order to avoid being paralyzed by the ensuing regress, it is important to identify the level of certainty that is acceptable, and that is likely to be acceptable to the performer as well.

More generally speaking, the expansion to the context of *joint* music perception and performance also raises the possibility of exploring new links between music perception and performance, on the one hand, and research on social cognition and social interaction on the other. There is a wealth of ongoing research, for example, on the (conscious or unconscious) signaling that occurs within joint action: in some cases, agents reduce the variability of their movements in order to render them more easily predictable for their partners; in other cases, they exaggerate certain movements in order to make them more easily interpretable (Vesper, Butterfill, Sebanz & Knoblich, 2010; Vesper, van der Wel, Knoblich & Sebanz, 2013; Pezzulo and Dindo 2011). One highly valuable benefit of investigating music perception and performance in joint actions is that it is possible to draw upon this rich background of research on joint action in order to motivate specific hypotheses prior to conducting research, and thus to constrain and structure theorizing in advance. By the same token, researchers working in the field of joint action should be delighted to explore a relatively fresh domain in which existing hypotheses can be tested, and potentially corroborated, in novel contexts.

## FURTHER DIRECTIONS

In sum, Schaefer offers a commendably clear and balanced presentation of how the HPP framework may fruitfully be applied in research on music perception and performance, and in doing so raises the possibility of enriching this research area with new theoretical and empirical resources from further afield. In order to exploit the potential benefits of this enrichment, it will be important to work towards the specification of hypotheses that are uniquely generated by the HPP framework. If this can be achieved, and if some or many of those hypotheses are borne out, it would not only be a boon to research on music perception and performance, but potentially also to research on joint action more generally, and may help to provide further empirical support to the broader HPP framework.

## NOTES

[1] Correspondence can be addressed to: Dr. John Michael, Department of Cognitive Science, Central European University, Frankel Leo Ut 30-34, 1023 Budapest, Hungary, michaelj@ceu.hu

## REFERENCES

- Barron, F. X. (1955). The disposition toward originality. *Journal of Abnormal and Social Psychology*, 51, 478–485.
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. East Norwalk, CT: Appleton-Century-Crofts.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204
- De Bruin, L., & Michael, J. (under review) Social cognition: A Bayesian predictive coding approach.
- Francès, R. (1976). Comparative effects of six collative variables on interest and preference in adults of different educational levels. *Journal of Personality and Social Psychology*, 33, 62–79.
- Friston, K. (2012). A free energy principle for biological systems. *Entropy*, 14, 2100–2121.
- Friston, K & Frith, C. (in press). A duet for one. *Consciousness and Cognition*.  
doi:10.1016/j.concog.2014.12.003
- Geeves, A., McIlwain, D., Sutton, J. & Christensen, W. (2013). To think or not to think: The apparent paradox of expert skill in music performance. *Educational Philosophy and Theory*, 46(6), 674–691.
- Gergely, G. and Watson, J.S. (1999). Early social-emotional development: contingency perception and the social biofeedback model. In P. Rochat (Ed.) *Early Social Cognition* (pp. 101–137). Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.
- Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.
- Hohwy, J., & Palmer, C. (2014). Social cognition as causal inference: implications for common knowledge and autism. In M. Gallotti and J. Michael (Eds.), *Social ontology and social cognition*. Volume 4 in the Springer Series “Studies in the Philosophy of Sociality”, Dordrecht: Springer.
- Hohwy, J., Roepstorff, A., Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687–701.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.
- Kammann, R. (1966). Verbal complexity and preferences in poetry. *Journal of Verbal Learning and Verbal Behavior*, 5, 536–540.
- Keller, P.E., Novembre, G., & Hove, M.J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B*, 369, 20130394.
- Michael, J., Christensen, W. & Overgaard, S. (2014). Mindreading as social expertise, *Synthese* 191, 817–840.
- Pezzulo, G., & Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research*, 211, 613–630.

- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6), 1236-1256.
- Reber, R., Schwarz, N. and Winkielman, P. (2004) Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4): 364-382.
- Repp, B. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*. 12: 969–92.
- Schaefer, R. (2014) Musical representations in musical processing and their role in action-perception loops. *Empirical Musicology Review*, 9(3/4): 161-176.
- Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge, MA: Cambridge University Press.
- Vesper, C., S. Butterfill, N. Sebanz, and G. Knoblich. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8/9), 998–1003.
- Vesper, C., van der Wel, P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 48-61.
- Wolf, T. Vesper, C., Knoblich, G. and Sebanz N. (2014). Do you believe in Mozart? The influence of beliefs on joint action outcomes. Poster presented at the *12th Biannual Conference of the German Cognitive Science Society*, 29 September – 2 October 2014.